DOCUMENT RESUME

ED 437 247 RC 022 200

AUTHOR	Filmer, Deon; Pritchett, Lant
TITLE	Estimating Wealth Effects without Expenditure Data--or Tears: An Application to Educational Enrollments in States of India. Policy Research Working Papers No. 1994.
INSTITUTION	World Bank, Washington, DC.
PUB DATE	1998-00-00
NOTE	39p.
CONTRACT	RPO-682-11
AVAILABLE FROM	World Bank, 1818 H Street, NW, Room MC3-638, Washington, DC 20433. Web site: <http://www.worldbank.org/research/>.
PUB TYPE	Reports - Research (143)
EDRS PRICE	EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Economic Status; *Educational Attainment; *Enrollment; *Family Financial Resources; *Family Status; Foreign Countries; *Measurement Techniques; National Surveys; Rural Education; Socioeconomic Influences; Validity
IDENTIFIERS	*India

ABSTRACT

The relationship between household wealth and educational enrollment of children can be estimated without expenditure data. A method for doing this uses an index based on household asset ownership indicators. To estimate the relationship between household wealth in India and the probability that a child aged 6-14 would be enrolled in school, data were drawn from the National Family Health Survey. However, this survey measured neither household income nor consumption expenditures. As a proxy for long-run household wealth, a linear "asset index" was constructed from asset indicators in the survey, using principal components analysis to derive the weights. The analysis used rural-only data or controlled for rural-urban differences in infrastructure that might bias rural household assets. The asset index is robust, produces internally coherent results, and closely corresponds with state domestic products and poverty rates. It was validated using data on consumption spending and asset ownership from Indonesia, Nepal, and Pakistan. The asset index has reasonable coherence with consumption expenditures and works as well as, or better than, traditional expenditure-based measures in predicting enrollment status. On average, a child from a wealthy household (top 20 percent of the asset index) was 31 percent more likely to be enrolled in school than a child from a poor household (bottom 40 percent). Varying across states, the wealth gap was 5-43 percentage points for all areas and 4-53 points for rural areas. (Contains 15 references and 14 data tables.) (SV)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 437 247

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY:

J. Feather

2A

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Estimating Wealth Effects without Expenditure Data -- or Tears:

An Application to Educational Enrollments in States of India

Deon Filmer

Lant Pritchett

Abstract: We use the National Family Health Survey (NFHS) data collected in Indian states in 1992 and 1993 to estimate the relationship between household wealth and the probability a child (aged 6 to 14) is enrolled in school. A methodological difficulty to overcome is that the NFHS, modeled closely on the Demographic and Health Surveys (DHS), measures neither household income nor consumption expenditures. As a proxy for long-run household wealth we construct a linear index from a set of asset indicators using principal components analysis to derive the weights. This "asset index" is robust, produces internally coherent results, and provides a close correspondence with State Domestic Product (SDP) and poverty rates data. We validate the asset index using data from Indonesia, Pakistan and Nepal which contain data on both consumption expenditures and asset ownership. The asset index has reasonable coherence with current consumption expenditures and most importantly, works as well, or *better*, than traditional expenditure based measures in predicting enrollment status. When the asset index is applied to the Indian data the results show large, and variable, wealth gaps in the enrollment of children across states of India. While on average across India a rich (top 20 percent of the asset index) child is 31 percentage points more likely to be enrolled than a poor child (bottom 40 percent), this wealth gap varies from only 4.6 in Kerala, to 38.2 in Uttar Pradesh and 42.6 percentage points in Bihar.

The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank, its Executive Directors, or the countries they represent.

RC 022200

BEST COPY AVAILABLE

Estimating Wealth Effects without Expenditure Data -- or Tears: An Application to Educational Enrollments in States of India[1]

Introduction

This paper has an empirical and overtly methodological goal. We propose and defend a method for estimating the effect of household economic status on educational outcomes without direct survey information on income or expenditures. We construct an index based on indicators of household assets, solving the vexing problem of choosing the appropriate weights by allowing them to be determined by the statistical procedure of principal components. While the data for India cannot be used to compare alternative approaches we use data from Indonesia, Nepal, and Pakistan which have both expenditures and asset variables for the same households. With these data we show that not only is there a correspondence between a classification of households based on the asset index and consumption expenditures but also that the evidence is consistent with the asset index being a *better* proxy for predicting enrollments--apparently less subject to measurement error for this purpose--than consumption expenditures.

This methodology of constructing an index of household economic status based on an asset index built from weights chosen by principal components is of potentially broad application. Nearly identical DHS (and NFHS) surveys have been carried out in

[1] We would like to thank Harold Alderman, Zoubida Allaoua, Gunnar Eskeland, Jeffrey Hammer, Keith Hinchliffe, Valerie Kozel, Alan Krueger, Peter Lanjouw, Marlaine Lockheed, Berk Ozler, and Martin Ravallion for valuable comments and discussions. This research was funded in part through a World Bank research support grant (RPO 682-11).

over 35 countries with information on assets but not consumption expenditures.[2] A consistent method for estimating household wealth from these surveys allows comparisons across countries in the wealth gaps between countries for a range of socio-economic outcomes. A companion paper uses the asset index to examine wealth gaps in educational attainment in 35 countries (Filmer and Pritchett, 1998a). This method allows an examination of differences in health outcomes and health care utilization across wealth groups from DHS data (Hammer, 1998). In addition, this method can be applied in studies of fertility and family planning usage. Beyond its use to estimate wealth effects on outcomes, the index is also a convenient control for household economic status. When examining the effects of other factors (such as maternal education on child health) one needs to control for household economic status and the method proposed provides a simple technique for doing so.

I) The NFHS data: Creating an Index

The National Family Health Surveys (NFHS) data, collected in 1992 and 1993 present both an opportunity, as well as a challenge. The opportunity is a set of surveys that followed nearly identical questionnaires for each Indian state with large samples designed to be representative at the state level.[3] The number of households surveyed in each state varied from 9,963 in Uttar Pradesh to around 1,000 in the small northeastern states. Overall, the survey covered over 88,000 households and about half a million individuals.

---

[2] With the exception of the 1994 Indonesian DHS.

An educational history was obtained for each household member.[4] For each member the survey asked whether they had ever been to school, the highest grade attended, and, for members less than 15 years old, if they were still in school.

The major challenge is that a household's economic status is undoubtedly an important determinant of enrollment, yet the NFHS did not collect information on either household income or consumption expenditures. However, the NFHS did inquire about household ownership of various assets and characteristics of the household's dwelling. We use twenty-one of these asset variables, which can be grouped into three types. First, eight variables about household ownership of certain consumer durables (clock/watch, bicycle, radio, television, bicycle, sewing machine, refrigerator, car). Second, twelve variables describing characteristics of the household's dwelling (three about toilet facilities, three about the source of drinking water, two about rooms in the dwelling, two about the building materials used, and one each about the main source of lighting and cooking). Third, a variable about whether the household owned more than 6 hectares of land.

These variables can be used to create an index of assets that proxies for household "wealth" or economic status. We limit our problem in this paper to forming a linear index of these asset variables to use as a proxy for household wealth in explaining school enrollment. Even within these limits choosing weights is a hard problem. There are three solutions that have been used in the literature. First, equal weights of all the assets, which

---

[3] The NFHS surveys were modeled closely on the Demographic and Health Surveys (DHS) of which there have been almost a hundred carried out in over 50 developing countries over the past two decades.

[4] Households without an woman eligible for the female questionnaire are still included in the part of the household survey which includes the education questions.

has as its *only* appeal not seeming as completely arbitrary as it really is. The second possible solution is to impose a set of weights. For instance, prices of various assets could be used to construct an index of household wealth, but this is possible only if the prices of various assets are available.[5]

A third solution is to not construct an index but simply enter all asset variables individually in a multivariate regression equation. This is the approach recommended in Montgomery, Burke, Paredes and Zaidi (1997) for use in fertility or mortality regressions using DHS data. This approach does handle the problem of "controlling" for wealth in estimating the impact of non-wealth variables. However, as recognized by Montgomery *et al* (1997), it does not identify the wealth effect as many assets play a both a direct and an indirect effect on outcomes. For instance, the household's use of electricity for lighting may have a role both as a proxy for wealth but also have the effect of making study easier and hence lower the opportunity costs of schooling. Or the availability of piped water may both indicate greater wealth but also reduce water collection times and lower the relative cost of schooling. There is no way to infer from the unconstrained coefficients on the asset variables from a multivariate regression the impact of an increase in wealth. Hence, while in some sense a regression coefficients produce a linear "index" of the asset variables (that which best predicts the dependent variable) this "index" cannot be interpreted as the effect of an increase in wealth.

---

[5] While this is a desirable solution and is done as part of estimating total consumption expenditures in surveys such as the LSMS, asset prices were not collected in connection with the DHS or NFHS and hence is impossible here.

A) *Using principal components*

We implement a different approach: we use the statistical procedure of principal components (which is closely related to factor analysis) to determine the weights for an index of the asset variables. Intuitively, principal components is technique for extracting from a large number of variables those few orthogonal linear combinations of the variables that best capture the common information. The first principal component is the linear index of variables with the largest amount of information common to all of the variables.

The result of principal components is an asset index is for each household *(Aj)* based on the formula:

$$A_j = f_1 \times (a_{j1} - a_1) / (s_1) + ... + f_N \times (a_{jN} - a_N) / (s_N)$$

where f_1 is the "scoring factor" for the first asset as determined by the procedure, a_{j1} is the *j* th household's value for the first asset and a_1 and s_1 are the mean and standard deviation of the first asset variable over all households. Our crucial assumption, and it is just that, an assumption, is that household long-run wealth is what causes the most common variation in asset variables.

The mean value of the index is zero by construction. The standard deviation in this case is 2.3. Since all the asset variables (except "number of rooms") take only the values of zero or one, the weights have an easy interpretation. A move from 0 to 1 changes the index by f_i/s_i. A household that owns a clock has an asset index higher by .54 than one that does not. Owning a car raises a household's asset index by 1.21 units. Using biomass for cooking lowers the index by .67.

Using this index each household is assigned to the bottom 40 percent, the middle

40 percent, or top 20 percent of households in all of India.[6] Purely for expository convenience, we will refer to these as the poor, the middle and the rich, asking the reader to keep firmly in mind that this is not following any of the usual definitions of "poverty" and that we are not proposing the asset index for use in poverty analysis or as a proxy for *current* living standards.

Table 1: Scoring factors and summary statistics for variables entering the computation of the first principal component

		All India			Mean		
	Scoring factors	Mean	Std. Dev.	Scoring factor * Std. Dev.	Poorest 40 percent	Middle 40 percent	Richest 20 percent
Own clock/watch	0.270	0.533	0.499	0.54	0.164	0.739	0.985
Own bicycle	0.130	0.423	0.494	0.26	0.264	0.510	0.621
Own radio	0.248	0.396	0.489	0.51	0.101	0.522	0.838
Own television	0.339	0.209	0.407	0.83	0.000	0.127	0.866
Own sewing machine	0.253	0.182	0.385	0.66	0.015	0.179	0.580
Own motorcycle/scooter	0.249	0.082	0.274	0.91	0.001	0.031	0.375
Own refrigerator	0.261	0.068	0.252	1.04	0.000	0.006	0.353
Own car	0.129	0.012	0.107	1.21	0.000	0.001	0.059
Drinking water from pump/well	-0.192	0.609	0.488	-0.39	0.800	0.569	0.242
Drinking water from open source	-0.041	0.040	0.195	-0.21	0.057	0.036	0.005
Drinking water from other (non-piped) srce	-0.002	0.019	0.138	-0.01	0.016	0.027	0.012
Flush toilet	0.308	0.217	0.412	0.75	0.005	0.175	0.797
pit toilet/latrine	0.040	0.086	0.280	0.14	0.040	0.127	0.111
none/other toilet	0.001	0.001	0.029	0.03	0.001	0.001	0.001
main source of lighting electric	0.284	0.510	0.500	0.57	0.143	0.700	0.989
Number of rooms in dwelling	0.159	2.676	1.957	0.08	1.975	2.965	3.739
kitchen is a separate room	0.183	0.536	0.499	0.37	0.312	0.643	0.848
main cooking fuel is wood/dung/coal	-0.281	0.776	0.417	-0.67	0.956	0.841	0.224
dwelling all high quality materials	0.309	0.237	0.425	0.73	0.005	0.218	0.821
dwelling all low quality materials	-0.273	0.483	0.500	-0.55	0.832	0.308	0.017
own >6 acres land	0.031	0.115	0.319	0.10	0.075	0.155	0.126
Economic status index		0.000	2.32		-2.00	0.071	3.857

Note: Each variable besides number of rooms takes the value 1 if true, 0 otherwise. Scoring factor is the "weight' assigned to each variable (normalized by its mean and standard deviation) in the linear combination of the variables that constitute the first principal component. Source: Authors' calculations from NFHS 1992-93

The difference in the average index between the poor and middle is 2.07 units. One example of a combination of assets that would produce this difference is owning a radio (.54), having a kitchen as a separate room (.37), having electricity for lighting (.57),

[6] Cutoff points for these quantiles were based on a ranking of individuals, that is the bottom 40 percent refers to the households in which the bottom 40 percent of people live.

BEST COPY AVAILABLE

and having a dwelling not of all low quality materials (-.55). The richest 20 percent have a wealth index almost four units higher than the middle 40. This additional difference is equivalent to owning a motor scooter (.91), a television (.83), having a flush toilet (.75), a house of all high quality materials (.73) and not using biomass as a cooking fuel (.67).

B) The reliability of the asset index

The asset index for India does well in three dimensions: first, it is internally coherent and produces clean separations across the poor, middle and rich households for each asset individually, second, it is robust to the assets included, third, it produces reasonable comparisons with poverty and output across states. However, the index does have its drawbacks, especially problems with urban/rural comparisons.

Internal coherence. The last three columns of Table 1 compare the average asset ownership across the poor, middle and rich households. The index produces sharp differences across groups in nearly every asset: clock ownership is 16 percent for the poor versus 98 percent for the rich, while the poor use biomass (wood/dung/coal) almost exclusively (96 percent) only 22 percent of the rich do so. One question is whether the asset index loads excessively on variables that are dependent on locally available infrastructure (electricity, piped water) rather than household specific variables. On this score the clean separation between poor and rich on non-infrastructure variables, for example, "all high quality materials in the dwelling" (only .5 percent of the poor versus 82.1 percent of the rich) and having a kitchen as a separate room (31 percent of the poor versus 85 percent of the rich) is reassuring.

Robustness. The asset index produces very similar classifications when different

subsets of variables are used in its construction. Table 2 reports the fraction of households classified in the bottom 40 when using all assets compared with indices based on (1) only the ownership of assets (watch, radio, etc. ...) (2) the ownership of assets, housing quality and number of rooms, and land ownership and (3) all the variables except those related to drinking water and toilet facilities. Almost no households classified in the poorest group by the index using all variables would be classified as "rich" by any of the more limited measures. The robustness of the classification is similar for the middle and rich groups.

Table 2: Classification differences of the bottom 40 percent by asset index constructed from different sets of variables: All India

		Only 8 asset ownership variables	Asset ownership, housing, and land ownership	All variables except for drinking water and toilet facilities
Groups based on asset index using all variables	Bottom 40 pct.	80.24	87.72	95.08
	Middle 40 pct.	19.70	12.28	4.92
	Top 20 pct.	0.06	0.00	0.00
	Total	100.00	100.00	100.00

Source: Authors' calculations from NFHS, 1992-93

Comparisons across states. Since the poor, middle, and rich are defined on an all India basis, states differ in the number of households in each group and hence we can compare state by state rankings with conventional measures. Nationwide, the expenditures poverty rate was 36 percent and hence is roughly comparable to the fraction "asset poor" in the bottom forty percent by the asset index. The first and second columns of Table 3 show the two classifications agree that Punjab, Haryana, and Kerala have better than average economic status and that Bihar, Orissa and Uttar Pradesh are worse than average. The rank correlation of the poverty rate and the proportion asset poor is .794 (p-value<.001). There are differences: Maharashtra looks richer (27 percent asset poor versus 37 percent poverty rate) and Andhra Pradesh looks poorer (39 percent asset

poor, but poverty rate of only 22 percent).

Table 3: Distribution of individuals across groups and state level poverty and net domestic product (sorted by the percentage in the bottom 40 percent)

	Proportion Asset Poor	State poverty rate (headcount index)	Per capita net state domestic product
	Bottom 40 pct.		
Delhi	1.3		
Goa	5.6		10128
Himachal Pradesh	6.8	28.58	
Punjab	8.4	11.46	10857
Haryana	10.5	25.22	9609
Jammu	14.5		
Kerala	15.1	25.12	5065
Mizoram	18.1		
Nagaland	20.3		
Gujarat	26.8	24.15	7586
Maharashtra	26.9	36.82	9270
Karnataka	27.6	32.91	6313
Manipur	27.6		
Tamil Nadu	32.5	35.40	6205
Meghalaya	37.9		5769
Arunachal Pradesh	38.1		6359
Andhra Pradesh	39.0	21.87	5802
Rajasthan	39.7	27.46	5035
Tripura	41.8		
West Bengal	44.3	36.94	5901
Uttar Pradesh	48.6	41.55	4280
Madhya Pradesh	49.4	42.46	4725
Orissa	54.4	48.64	3963
Assam	58.3	41.09	5056
Bihar	61.5	55.15	3280
All India	40.0	36.16	6380

Notes: The rank correlation coefficient between the percent asset poor and the poverty rate is 0.794 (p-value <.001), the rank correlation between the percent asset poor and per capita state product is -0.864 (p-value <.001). Sources: NFHS, 1992/93 and Haque, Lanjouw and Ravallion, 1998, and Agrawal and Varma, 1996. Data on the Headcount Index are for 1993/94.

The rank correlation of the proportion asset poor and SDP per capita, is -.864 (p-value<.001).[7] While the rankings agree overall, again certain states look different by the two rankings. For example, Kerala looks richer by the index (only 15 percent asset poor with per capita SDP of 5065) while Assam looks poorer (58 percent asset poor) versus per capita SDP of 5056. However, the conventionally defined poverty rate agrees with assets against SDP showing Assam poorer than in Kerala (41 versus 25 percent).

BEST COPY AVAILABLE

On the other hand. The first principal component explains 25.6 percent of the variation in the twenty-one wealth variables, which is substantial, but not overwhelming. While the first principal component of assets might well serve as a reasonable overall index, a remaining question is whether the first component contains all of the relevant information. The second principal component is more difficult to interpret, but it appears to be capturing rich rural households.

This is particularly worrisome because the rankings by the asset index show rural households to be less "wealthy" than do conventional poverty measures. One explanation for this discrepancy is that since many of the asset variables depend on the availability of infrastructure (electricity, piped water, sewerage), urban households are more likely to appear well-off than poorer households. On the other hand this may well imply that standard poverty measures underestimate the difference between rural and urban households by not adjusting real incomes for the implicit price differentials for services provided by infrastructure. But back on the other hand, for the analysis of enrollment decisions we want an index that captures the dimensions of wealth relevant to education. Finally, in this particular application we abandon all hands, mix metaphors, and punt: the analysis below either uses rural only data or controls for rural/urban status so any level difference due to systematic over or under statement of the differences should not affect the analysis.

II) Asset index versus consumption expenditure as proxies for long-run wealth

Before using the asset index in an examination of the India educational data we

---

[7] The rank correlation between the poverty rate and per-capita state domestic product is -.729 (p-value =.002)

make a methodological detour and ask how the results are likely to compare to those using more conventional rankings, such as by consumption expenditures. In making the comparison of an asset index and current expenditures we do not mean to imply that we are creating an asset index intended to serve as a proxy for expenditures.[8] Rather, often *both* are proxies for something unobserved: a household's long-run "wealth" or more broadly "economic status". Therefore, while it is reassuring that the two are related, discrepancies in the classification of households cannot be assumed to be "mistakes" of the asset index as they could just as easily be indicating limitations of current consumption. The two measures have conceptually distinct limitations. The problem with the asset index is not having appropriate weights for the assets. In contrast, the problem with *current* expenditures (as a proxy) is that it would only be a perfect measure for long-run wealth under the patently unrealistic assumption of perfect foresight and perfect capital markets.[9]

We address the comparison in two ways. First, we use household survey data from three countries and construct both an asset index and a consumption expenditures

---

[8] It is in making this distinction that our approach most differs from Montgomery *et al* (1997) which is the most comprehensive treatment of the issue of using asset variables in the DHS to date. In their work the issue is framed as an attempt to use the asset variables, or an asset index created from them, as a proxy for per-capita consumption. That is, the quality of any measure, or measures, used is assessed from a diagnostic regression of consumption on the asset measure(s). Moreover, their discussion on the effect of asset variables on an outcome measure (e.g. child mortality) is about what this means for inferences about the effect of consumption expenditures on the outcome.

[9] The main reason why current expenditures is a popular proxy for long-run wealth are both the theoretical justification that expenditures are superior to current income as a proxy for long-run income because of consumption smoothing and, perhaps even more important, the pragmatic justification that expenditures are easier to measure than income in most rural settings. On neither of these dimensions is current expenditures unambiguously preferred over asset ownership.

BEST COPY AVAILABLE

based measure for the same households and compare classifications based on the two.[10] Second, we compare the relationship between enrollment rates and wealth using the asset index and expenditures.

A) Comparisons of consumption expenditures and asset index classifications

The Nepal Living Standards Survey carried out in 1996 (NLSS) and the Pakistan Integrated Household Survey carried out in 1991 (PIHS) are "standard" Living Standards Measurement Study (LSMS) Surveys (Grosh and Glewwe, 1998). The Indonesian DHS carried out in 1994 (IDHS) included an experimental consumption expenditures module, based closely on Indonesia's SUSENAS survey, for about half of the households.

For each country we constructed a principal components assets index as well as used (or derived) a measure of household size adjusted consumption expenditures.[11,12] Individuals can be assigned to percentile based groups (bottom 40, middle 40, top 20) using either the asset index or the expenditure measure. Table 4 shows the results of comparing the two classifications. The results in Indonesia and Nepal are quite similar. Roughly two-thirds of those classified into the bottom 40 by expenditures are also classified into the bottom 40 by assets and only 5 percent of those in the bottom 40

---

[10] We compare the asset index to consumption expenditures and not *predicted* consumption expenditures, where assets and other household variables are used as instruments. While some of the results would appear more similar if we had taken this "best practice" approach (recently used and explored in Behrman and Knowles, 1997) the conventional approach—particularly for bivariate/tabular analysis—is to not use predicted expenditures and therefore we use this as our baseline for comparison.

[11] The weights on the various assets are reassuringly similar with indices for the three countries: e.g. large negative weights on using biomass fuels, large weights on owning a motorcycle or scooter, and large weights on quality of housing materials.

[12] We used total consumption expenditures (C) adjusted for household size (N), C/N^{α}, where the adjustment for economies of scale α equals 0.6 (see Lanjouw and Ravallion, 1995, and Drèze and Srinivasan, 1997, for discussions of this parameter).

percent by expenditures appear in the top 20 percent by assets. The classification of the richest 20 shows less agreement; between 49 and 56 percent of those rich by expenditures are also in the top 20 by assets, but 10 to 13 percent of those ranked in the top 20 by expenditures are in the bottom 40 by assets.

Table 4: Classification differences using asset index based groups and groups derived from household consumption expenditures in Nepal, Indonesia, and Pakistan

		Groups based on household consumption per adjusted size*	
Nepal			
		Bottom 40 pct.	Top 20 pct.
Groups based on asset index	Bottom 40 pct.	65.20	12.63
	Middle 40 pct.	29.85	31.41
	Top 20 pct.	4.95	55.96
	Total	100.00	100.00
Indonesia			
		Bottom 40 pct.	Top 20 pct.
Groups based on asset index	Bottom 40 pct.	63.91	10.43
	Middle 40 pct.	31.58	41.06
	Top 20 pct.	4.51	48.50
	Total	100.00	100.00
Pakistan			
		Bottom 40 pct.	Top 20 pct.
Groups based on asset index	Bottom 40 pct.	60.48	21.77
	Middle 40 pct.	35.15	35.52
	Top 20 pct.	4.37	42.71
	Total	100.00	100.00

* Adjusted household size is equal to household size to the power 0.6.
Source: Authors' calculations from NLSS 1996, IDHS, 1994, and PIHS, 1991

The results for Pakistan show less coherence between the two rankings. While it is still the case that only 4 percent of those that are poor by expenditures are rich by assets, only 60 percent of the expenditure poor are also asset poor. Moreover, only 43 percent of those in the top 20 by expenditures are also in the top 20 by assets and 22 percent of the *top* 20 percent of households by expenditures are in the *bottom* 40 percent by assets.[13]

[13] Many of the assets, like the quality of materials, are at the household level and benefit all household members so our asset index is unadjusted for household size. Generally the fit between assets and expenditure classifications is better the smaller α, so that the asset index classification fits total

B) Comparison of enrollment rates using the two measures

The main purpose of classifying households by economic status is to examine what fraction of children in each wealth group is in school. Table 5 compares differences in enrollment and completion indicators between the rich and poor groups when the calculation is done based on either the asset index or expenditures.

Table 5: Difference between the average for the highest 20 percent and the bottom 40 percent of the outcome indicators using asset index and consumption expenditures: Nepal, Indonesia and Pakistan

Percentage point difference between the top 20 and bottom 40 percent in:	Based on Asset index	Based on household consumption expenditures (adjusted for size)	Difference between the asset index and the expenditures classification
Nepal			
Ever went to school (ages 6 to 14)	41	40	1
Currently attending school (ages 6 to 14)	42	41	1
Completed at least grade 6 (ages 15 to 19)	48	49	-1
Indonesia			
Ever went to school (ages 6 to 14)	10	7	3
Currently attending school (ages 6 to 14)	19	15	4
Completed at least grade 6 (ages 15 to 19)	23	16	7
Pakistan			
Ever went to school (ages 6 to 14)	34	26	8
Currently attending school (ages 6 to 14)	33	26	7
Completed at least grade 6 (ages 15 to 19)	44	33	11

Source: Author's calculations from IDHS 1994, NLSS 1996, PIHS, 1991

In Nepal the results are almost identical: the proportion of 6 to 14 year olds from rich households who ever attended is 41 percentage points higher than poor households when defined by assets, and 40 percentage points higher for the rich than poor when defined by expenditures. In Pakistan in contrast, the gap in the proportion of 6 to 14 year olds who ever attended school is 34 percentage points when rich and poor are defined

household expenditures better than what is reported and fits per capita expenditures worse than is reported.

according to assets while there is a smaller rich-poor gap, 26 percentage points, when the groups are defined on expenditures. The Indonesian results are in between: the wealth gap in current enrollment is 19 percentage points based on assets but only 15 percentage points based on expenditures. For all three countries classifying households by the asset index consistently to produces a larger gap between the rich and the poor than classifying households by expenditures.

Table 6: Difference in average enrollment rates between the richest 20 percent and the poorest 20 percent using asset index and consumption based measures to derive quintiles.

	Difference between the enrollment rates of rural children aged 6-14 from the top and bottom quintiles when household quintiles are constructed by:		Difference between the asset index and consumption based Classification
	Asset index	Per capita consumption expenditures	
Andhra Pradesh	55	37	19
Assam	36	21	15
Bihar	67	43	25
Gujarat	46	27	19
Haryana	49	39	10
Karnataka	51	38	13
Kerala	12	3	9
Madhya Pradesh	55	33	22
Maharashtra	34	25	9
Orissa	47	38	10
Punjab	56	52	5
Rajasthan	52	41	11
Tamil Nadu	25	15	9
Uttar Pradesh	52	30	21
West Bengal	51	40	11

Source: Author's calculations from NFHS, 1992-93. Enrollment for consumption quintiles from Haque, Lanjouw and Ravallion, 1998

While we cannot compare the same households using Indian data, we can compare averages for the NFHS data and the asset index with averages for Indian National Sample Survey (NSS) data and per capita consumption expenditures (i.e. $\alpha=1$, from Haque, Lanjouw and Ravallion (1998). Table 6 reports, for rural areas of each state, the difference in the enrollment rate between the top and bottom quintiles. The "flatter" wealth-education profile found for rankings based on the consumption expenditures is

BEST COPY AVAILABLE

true for *every* state of India for which we can make the comparison.

C) Measurement error in proxies for long-run wealth

The first two columns of Table 7 show, by quintile, the fraction of children aged 6 to 14 in rural areas of India enrolled when children are classified by the asset index from the NFHS data or by per capita consumption expenditures from the NSS data (Haque, Lanjouw and Ravallion, 1998). While the enrollment rates for the middle quintile from the two sources agree almost exactly (70 versus 71 percent), the enrollment rate profile based on quintiles from household consumption expenditures from NSS data has a "flatter" profile (from 49 to 82) than the profile based on an asset index. The enrollment of the poor is 7 percentage points lower (49 versus 42) using the assets index, while the enrollment of the rich is 12 percentage points higher (94 versus 82). Therefore the raw "wealth gap" in enrollment rates is 33 percentage points with consumption expenditures but 52 percentage points using the asset index.

The last three columns of Table 7 display a hypothetical calculation of the effect of measurement error on the wealth-enrollment profile. Using plausible values for the degree of measurement error of either measure as a proxy for long-run wealth either from assets or the magnitude of transitory component in consumption expenditures one can find substantial effects in "flattening" the enrollment/wealth profile. As is to be expected the larger measurement error in relation to the "true" variation of income the "flatter" the enrollment-income profile and the lower the differences between quintiles. Even if the "true" wealth-gap were 60 points then a modest amount of measurement error would reduce the observed gap to 51 and substantial measurement error would reduce the gap

further still.

Table 7: Enrollment rates by quintile, household per capita consumption and asset index and an illustration of the attenuation effects of measurement error.

	Enrollment of rural children aged 6-14 when household quintiles are constructed by:			Hypothetical enrollment profile when the noise to total (noise plus signal) ratio is:	
Quintile:	Per capita consumption expenditures	Asset index	Assumed true profile	20 percent	50 percent
1	49	42	40	45	50
2	61	58	55	57	61
3	70	71	70	70	70
4	76	84	85	83	79
5	82	94	100	96	90
Difference between quintiles 5 and 1	33	52	60	51	40

Source: Enrollment by consumption quintiles from Haque, Lanjouw and Ravallion, 1998, final columns from Monte Carlo simulations.

Some additional insight about the role of measurement error comes from a heuristic use of regression analysis.[14] We use two approaches to exploring the relative amounts of measurement error in the two variables: instrumental variables (IV) and reverse regression. Under the hypothesis that expenditures and the asset index are both proxies for long-run wealth and that the measurement error of each is not perfectly correlated then each proxy can be used as an instrument for the other to mitigate the measurement error attenuation bias. The ratio of OLS to IV estimates is an estimate of the relative signal to signal plus noise for the two variables. The lower the ratio the worse the variable is as a proxy for predicting enrollments. This is true even if the measurement error in expenditures and assets is correlated and hence neither of the IV estimates is consistent. The degree of inconsistency in the IV estimates depends only on the

measurement error common to both measures and hence IV estimates for both expenditures and assets will converge to the same number. In contrast, the degree of inconsistency in the OLS depends on both the common and the indicator specific measurement error. Hence the ratio of the ratio of OLS to IV for each measure is a valid indicator of the *relative* degree of measurement error.

In Nepal, when we regress current enrollment on the asset index using consumption as an instrument, the ratio of the OLS to IV estimates is 0.66, while when we regress enrollment on the consumption measure using the asset index as an instrument, the ratio of the OLS to IV estimates is 0.46, yielding a ratio of the two of 1.4 (Table 8).[15] In Indonesia and Pakistan, when we regress current enrollment on the asset index and use consumption expenditures as an instrument, the ratio of the OLS to IV estimate is 0.85 in Indonesia and 1.00 in Pakistan. When we regress enrollment on the consumption measure using the asset index as an instrument, the ratio of the OLS to IV estimates is 0.16 in Indonesia and 0.15 in Pakistan, yielding a ratio of the ratios of 5.3 and 6.7 respectively.[16]

---

[14] We call these regressions "heuristic" as we are estimating extremely simplified linear probability models including only dummy variables for urban residence and male gender as controls to examine the issue of measurement error.

[15] Behrman and Knowles (1997) find that their estimates of the elasticities of various education outcome measures with respect to income, or consumption expenditures, in Vietnam increase by between 50 to 60 percent when they use household assets and other household characteristics as instruments for consumption. In the India case we do not have expenditures, but we do have 21 assets. So we divided the asset variables into two groups and constructed an asset index out of each set to form repeat measurements on long-run wealth. While both of these will be imperfect proxies for long-run wealth, the measurement errors will not be perfectly correlated and hence each can be used as an instrument for the other. In this case the ratio of OLS to IV estimates of around 1/2 is an estimate of the variance of the "true" to the total variance. The wealth index appears to have a substantial measurement error component.

[16] The would indicate an extraordinary degree of measurement error in Indonesia and Pakistan's consumption expenditure data. In Pakistan this is consistent with the fact that the R-squared of regressing consumption expenditures on the assets in Montgomery *et al* (1997) is only .13 in Pakistan

An alternative approach to measurement error is to use reverse regression for the bivariate relationships. That is, regress enrollment on the wealth measure and estimate the coefficient on wealth (β), then regress the wealth measure on enrollment and estimate the coefficient on enrollment (δ). If enrollment and wealth are measured with error then the true regression parameter is bounded by β (which is biased towards zero because of attenuation bias) and $1/\delta$ (which is biased away from zero as δ is biased towards zero because of attenuation bias).

A comparison of the ratio of β and $1/\delta$ when using the asset index and when using expenditures again estimates the *relative* measurement error in the two variables (as whatever measurement error is in enrollment is the same for the two analyses). In Nepal the reverse regression yields an estimate which is 12 times higher than the direct regression when using the asset index and 15 times higher when using expenditures, yielding a ratio of 1.3 (Table 8). This is very close to the ratio of 1.4 from the IV approach. In Indonesia the reverse regression estimate is 20 times higher for the asset index and 56 times higher for consumption, and for Pakistan the numbers are 15 and 111, yielding ratios of 2.9 and 7.2 respectively—again very comparable to the IV results.

The results in Table 8 are all the stronger when compared to those in Table 5 which showed that the gap in the probability of enrollment between the bottom 40 percent and the top 20 percent was larger when households are ranked by the asset index, and that the difference was smallest in Nepal and largest in Pakistan. For current enrollment of children 6 to 14, the difference using the alternative ways of ranking households in the

---

versus .22 or above in three other countries they report (in addition to our result of .24 for Nepal). It is less consistent with the R-square being equal to .33 in Indonesian DHS.

.gap was 1 in Nepal, 4 in Indonesia, and 7 in Pakistan: not only is the ordering the same but the magnitudes are consistent.

Table 8: Enrollment as a function of the asset index or consumption expenditures: Alternative estimates of relative measurement error of expenditures versus asset index.

	IV method			Reverse regression method		
	OLS to IV ratio: Asset Index[a]	OLS to IV ratio: Cons. Expend.[b]	**Ratio of ratios**	Reverse to direct ratio: Asset Index[c]	Reverse to direct ratio: Cons. Expend.[d]	**Ratio of ratios**
Nepal	0.66	0.46	**1.4**	11.9	15.2	**1.3**
Indonesia	0.85	0.16	**5.3**	19.6	56.4	**2.9**
Pakistan	1.00	0.15	**6.7**	15.4	111.0	**7.2**

Note: (a) $(\beta^{AI}_{OLS}/\beta^{AI}_{IV})$ (b) $(\beta^{CE}_{OLS}/\beta^{CE}_{IV})$ (c) $(\beta^{AI}_{OLS}/\beta^{*AI}_{OLS})$ (d) $(\beta^{CE}_{OLS}/\beta^{*CE}_{OLS})$
Where β is the coefficient on the asset index or consumption in a regression of enrollment on the asset index or consumption, and β^* is the inverse of the coefficient on enrollment in the (reverse) regression of the asset index or consumption on enrollment. Source: Author's calculations from IDHS 1994, NLSS 1996, PIHS, 1991

D) Stability over time of household rankings

All of these results are consistent with much less "noise" in an asset index than in consumption expenditures—as a proxy for long-run wealth. We wish to stress that our discussion of "measurement error" in consumption expenditures needs to be understood not as a statement about error in the measurement of actual current consumption, rather in the measurement of an indicator for use as a determinant of educational outcomes. These are likely to be much less sensitive to transitory fluctuations in expenditures and therefore one explanation of the "superior" performance of the asset index is that household rankings based on this index are more stable than those based on a consumption measure.

A panel survey of households in Morocco from 1992 to 1995 provides the basis to explore this issue. A DHS survey in 1992 covered 6407 households and a 1995 survey covered 2751, of which 2489 households can be matched across surveys. Table 9 presents the classification differences across the two time periods for the subsample of

BEST COPY AVAILABLE

overlapping households.[17] For example, 78.4 percent of the households who are classified as being in the poorest quintile in 1992 are also in the poorest quintile in 1995, and essentially none (1.3 percent) move out of the bottom 40 percent.

Table 9: Classification differences using asset index derived from two samples (with overlap) in Morocco.

		Quintiles based on 1995 ranking					
		1	2	3	4	5	Total
Quintiles	1	78.4	20.3	1.1	0.2	0.0	100.0
based on	2	26.6	53.9	19.5	0.0	0.0	100.0
1992 ranking	3	2.8	24.5	54.7	13.6	4.5	100.0
	4	0.0	2.5	15.7	58.4	23.4	100.0
	5	0.0	0.0	3.1	32.6	64.2	100.0

Source: Authors' calculations from Morocco DHS, 1992 and 1995.

In a recent survey, Fields (1998) reports a similar analysis of stability of classifications based on expenditures for four countries. Table 10 summarizes the results on changes in household rankings from these studies. These results clearly show more variability over time for the income or consumption expenditure based classifications than the results for Morocco using the asset based measure, particularly for the poorest quintile. However, a major caveat here is that the Morocco panel spans only 3 years which is the shortest time span for all the countries compared.

Table 10: Stability over time in rankings, comparison from panel data sets.

Country	Start year	End year	Difference	Variable used to rank households (individuals)	Percent in the poorest quintile who stay in the poorest quintile	Percent in the richest quintile who stay in the richest quintile
Morocco	1992	1995	3	Household asset index	78	64
Malaysia	1967	1976	9	Income of males	55	62
Chile	1968	1986	18	Per capita household income	8	58
China (rural)	1978	1983	5	Household income	54	61

[17] Households are classified in each time period according to their position with respect to the entire sample not just the subsample that can be matched over time.

Table 10: Stability over time in rankings, comparison from panel data sets.

Country	Start year	End year	Difference	Variable used to rank households (individuals)	Percent in the poorest quintile who stay in the poorest quintile	Percent in the richest quintile who stay in the richest quintile
	1983	1989	6	Household income	41	49
Lima, Peru	1985	1990	5	P.C. household consumption	40	50

Source: Adapted from Fields (1998) and authors' calculations from Morocco DHS 1992 and 1995.

Based on a recent six year panel of households in China, Jalan and Ravallion (forthcoming) find that annual consumption expenditures have a high degree of variability. In particular, they find that the average standard deviation of consumption per person across households is 384 (the mean is 342 Yuan per person per year at 1985 prices over the period 1985-90) and that the mean of the intertemporal standard deviation for any given household, over the entire period, is 189. So the standard deviation of a household's measured expenditures over time is about half that in the cross-section which will imply substantial changes in the classification across years.

Methodological summary so far

In some ways we are standing the conventional wisdom exactly on its head. The conventional wisdom is that survey based household consumption expenditures are the best estimates, not only of current expenditures, but are also the best proxy for households long-run wealth, while surveys without consumption expenditures have limited value, as they cannot control for, or estimate, wealth impacts. However, there is no *a priori* argument as to why current consumption expenditures are a better proxy of long-run household economic status than an index of assets: it is an open empirical question. Our results suggest that a methodologically simple solution to the vexing problem of creating a weighted index, using the technique of principal components,

works. Rather then being an *ad hoc* embarrassment the asset index appears to be more stable, less contaminated with measurement error as a measure of long-run wealth, and hence predicts enrollment differences better than traditional consumption expenditures. This obviously has important implications for the uses of the DHS and NFHS data sets to examine a broad range of indicators.

III) Wealth gaps in educational outcomes in Indian states.

Armed with data on educational outcomes on the one hand and the newly constructed and (elaborately) defended proxy for wealth on the other, we now address how the enrollments and attainment of children differ within Indian states according to the economic status of the household, and controlling for that, how enrollment is affected by gender, location, and the presence of schools.

A) Descriptive statistics: raw wealth gap

Overall only 68 percent of children aged 6 to 10 and 66 percent of those aged 11 to 14 are reported as being in school.[18] As is well known, educational enrollments and attainments vary widely across Indian states. The percentage of 6 to 10 year olds in school ranges from only 50 percent in Bihar to 96 percent in Kerala, and the percentage of those 11 to 14 in school ranges from 54 percent in Bihar to 94 percent in Kerala and Mizoram. The percentage of adults 15 to 65 who have ever attended school ranges from

[18] This is dramatically less than would be suggested from official government enrollment data, a discrepancy explored extensively in a recent World Bank book on basic education in India (World Bank, 1997). The present analysis focuses on differences in enrollments across groups (wealth, gender) and therefore these differences in absolute levels are less relevant.

42 percent in Bihar to 93 percent in Mizoram, and has a national average of 55 percent. Average years of attainment of those who even attended to school, ranges much less than enrollments and is close to 8 years of schooling in all states (except Delhi).

Figure 1 shows the "attainment profiles" for those aged 15 to 19: each state specific graph shows the attainment of children who live in poor, middle, and rich households. The attainment profile shows the proportion of children who have completed any given grade or higher. The gap between 1 and the intercept shows the proportion who never enrolled (or more specifically, never completed grade 1) while the slope indicates the percentage of children who drop out across the years. Table 11 shows whether a child is enrolled and the probability a child aged 15 to 19 completed grade 8 classified by the household's asset index.[19]

Both Figure 1 and Table 11 show clearly that children from richer households do better in all states. On average 94 percent of children aged 6 to 14 from the upper 20 percent are in school. This high enrollment rate of the rich is remarkably consistent across states, it is above 90 percent in all but three states (Arunachal Pradesh, Assam, and Tripura). Moreover, in all states, when children from rich households enroll, they stay in school. The enrollment profiles for the richest group are virtually flat between grades 1 and 5 in all states, and only slightly decreasing between grades 5 and 8. This combination results in over 70 percent of 15 to 19 year olds from the richest economic group completing grade 8 in all but two states (Meghalaya, Arunachal Pradesh).

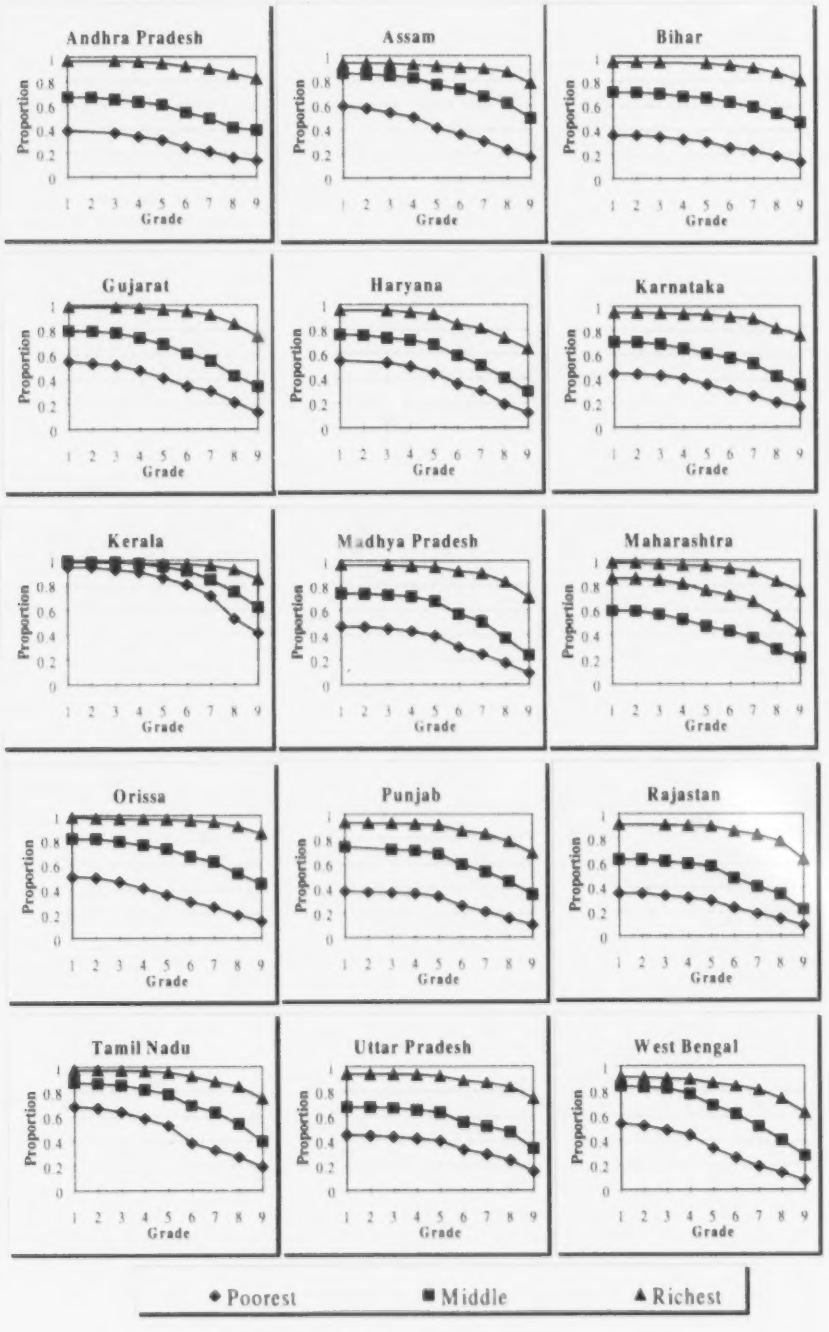
---

[19] Because the economic groups are based on the all India sample, there are sometimes very few observations from which to derive the numbers displayed here. When the number of observations for any subgroup drops below 40 the attainment profile is not shown. For example in Delhi there are very few observation in the lowest economic group and therefore the proportion is not reported for that group.

In sharp contrast, among the poorer part of the population educational attainment is dismal. Only half of the children aged 6 to 14 are in school. Moreover, the profiles suggest in several states (for example, Assam, Gujarat, Orissa, Maharashtra, Tamil Nadu, and West Bengal) that on top of a low proportion having completed grade 1, dropout is high leading to an even lower proportion who complete grade 5, and a substantially lower proportion for grade 8. Overall, only 38 percent of children aged 15 to 19 from the poor households finished grade 5. Only one in five poor children finished eight years of basic education.

The gap in educational enrollment and attainment between the rich and poor is enormous, but it also varies a great deal across states. The wealth gap in "ever enrolled" varies from a minor 9 percentage points in Kerala to a substantial 56 percentage points in Bihar. The gap in the attainment of grade 8 varies from (a non-negligible) 39 percentage points in Kerala to 72 percentage points in Orissa.

Figure 1: Attainment profiles for ages 15 to 19, by economic group



BEST COPY AVAILABLE

Table 11: Basic statistics on education status by wealth group (states sorted by average enrollment rate)

State	Proportion of 6 to 14 year olds who are currently "in school"				Proportion of 15 to 19 year olds who have completed at least grades 8			
	Average	Bottom 40 percent	Top 20 percent	Wealth gap (top - bottom)	All	Bottom 40 percent	Top 20 percent	Wealth gap (top-bottom)
Kerala	0.949	0.887	0.975	0.088	0.749	0.531	0.923	0.392
Goa	0.937	0.774	0.973	0.200	0.703	0.344	0.848	0.504
Himachal Pradesh	0.908	0.724	0.970	0.246	0.565	0.233	0.818	0.585
Mizoram	0.907	0.768	0.974	0.205	0.567	0.190	0.844	0.654
Manipur	0.902	0.804	0.991	0.186	0.610	0.359	0.927	0.568
Nagaland	0.896	0.824	0.980	0.157	0.572	0.354	0.865	0.511
Delhi	0.872	0.477	0.924	0.448	0.685		0.766	
Jammu	0.857	0.666	0.979	0.313	0.541	0.195	0.833	0.638
Tamil Nadu	0.825	0.717	0.950	0.232	0.518	0.269	0.838	0.570
Maharashtra	0.820	0.671	0.962	0.290	0.579	0.279	0.832	0.554
Haryana	0.813	0.605	0.957	0.352	0.480	0.189	0.728	0.539
Punjab	0.808	0.427	0.957	0.531	0.571	0.153	0.777	0.624
Tripura	0.795	0.710	0.873	0.163	0.395	0.187	0.789	0.603
Gujarat	0.757	0.552	0.962	0.410	0.504	0.212	0.845	0.633
Meghalaya	0.749	0.601	0.959	0.358	0.326	0.150	0.667	0.516
Arunachal Pradesh	0.711	0.585	0.865	0.279	0.340	0.184	0.585	0.400
Karnataka	0.708	0.507	0.943	0.437	0.447	0.205	0.816	0.611
Assam	0.703	0.615	0.846	0.231	0.422	0.229	0.866	0.637
Orissa	0.697	0.552	0.969	0.416	0.395	0.189	0.908	0.719
West Bengal	0.678	0.527	0.902	0.375	0.338	0.137	0.734	0.597
Andhra Pradesh	0.639	0.457	0.917	0.460	0.419	0.160	0.859	0.698
Madhya Pradesh	0.626	0.461	0.937	0.476	0.367	0.172	0.832	0.661
Uttar Pradesh	0.614	0.484	0.939	0.455	0.424	0.239	0.836	0.598
Rajasthan	0.593	0.414	0.91	0.496	0.345	0.141	0.773	0.632
Bihar	0.514	0.378	0.942	0.564	0.381	0.183	0.864	0.681
All India	0.677	0.500	0.942	0.442	0.447	0.204	0.824	0.620

Source: Calculated from NFHS data, 1992-93

The implication of the small differences among the rich and huge differences for the poor in enrollment and attainment across Indian states means that gaps are largely driven by the extent to which states have been able to reach the bottom part of the economic distribution and bring them into the educational system. For instance, Tamil Nadu and Rajasthan are not that different in the percent of the households asset poor: 37 percent in Tamil Nadu and 43 percent in Rajasthan. However, their average educational attainment is quite different: only 52 percent of 15 to 19 year olds completed grade 5 in

Rajasthan as compared to 74 percent in Tamil Nadu. What causes this large difference? In both states the attainment of grade 5 by the rich is high, 96 percent in Tamil Nadu versus 90 percent in Rajasthan. What differs is how likely the poor are to reach grade 5. While in Tamil Nadu 52 percent of the poor population reached grade 5, this was only true of 29 percent of the poor in Rajasthan, a gap between the two states of 23 percentage points.

B) Estimation of wealth effects with child, household, village and state controls

To disentangle the determinants of school enrollment, we now estimate a probit regression with the dependent variable of whether or not the child aged 6 to 14 is enrolled in school:

$$E_i^* = \sum_{j=2,5} \beta_j \times Q_{ij} + \alpha \times X_i + \sum_{k=2,25} \delta_k \times \lambda_{ik} + \varepsilon_i$$

The wealth effects are specified by including the Q_{ij} s which are dummy variables equal to one if child *i* is in quintile *j* (the reference quintile is the poorest group).

In all of the samples the variables included besides wealth (X_i) are: the *child* variables of a dummy variable for gender, child's age and age squared, the *household* variables of age of the head of the household, whether the household head ever attended school, the highest grade completed of the household head, whether the household is Hindu, whether the household is from a scheduled caste or tribe.[20] Last, the specification includes a set of dummy variables λ_{ik} equal to one if child *i* lives in state *k*.

The other variables present (in the set of X_is) depend on whether the sample

includes urban and rural households or is limited to rural areas (the data on school availability and other village characteristics is limited to rural areas). In the pooled urban/rural samples a dummy variable is included for urban location of the household. In the sample with rural areas only the variables include three dummy variables for the presence of (1) a primary school, (2) a primary and a "middle" school, and (3) a primary, "middle" and a secondary school. In addition a large set of other village level variables capturing village infrastructure is included (e.g. post-office, bank, cinema house).

Table 12 reports the estimation of the equation for the all India sample for the combined urban and rural and for the rural only samples. The results show that there is a strong wealth effect in the probability of enrollment. All else equal, a child from a household in the highest quintile is about 31 percentage points more likely to be in school than a child from the poorest quintile. Moreover, the effects are strictly ordered across the quintiles: being in the second quintile increases the probability of being in school by 10 percentage points and each subsequent quintile increases the probability by roughly 7 percentage points (10.3 to 16.9 to 24.1 to 30.7).

The results on wealth for rural areas only, where a host of additional village level factors are included in the model are included, are very similar. In particular, the rural sample includes information on school availability so these wealth effects represent the effects of wealth, even when controlling for the fact that the poor are more likely to live in villages without schools. Even with these additional controls the magnitude of the wealth effects are nearly identical to those in the all India sample (11.1, 18.5, 26.9, and

[20] If the information on the education of the head of the household is missing we set the head ever attended school and head's highest grade variables to zero and set an indicator dummy variable equal to one in the regression.

31.5).

Table 12: Marginal effects on the probability of being "in school" for ages 6 to 14, (Probit regression results)

		All India (urban and rural)		Rural only	
	Zero / one variable	Marginal Effect	T-ratio	Marginal Effect	T-ratio
Quintile 2[a]	*	0.103	12.32	0.111	9.87
Quintile 3	*	0.169	16.94	0.185	17.92
Quintile 4	*	0.241	22.55	0.269	20.77
Quintile 5	*	0.307	23.53	0.315	18.69
Male	*			0.237	8.42
Rural male[b]	*	0.070	3.85		
Urban Female	*	-0.107	-6.19		
Rural Female	*	-0.149	-6.70		
Scheduled caste / Scheduled tribe	*	-0.047	-3.87	-0.053	-4.37
Age		0.206	13.37	0.232	13.20
Age squared		-0.011	-16.89	-0.012	-16.47
Head is male	*	-0.092	-5.64	-0.119	-5.90
Head's age		0.001	4.29	0.002	5.41
Head ever attended school	*	0.072	6.73	0.071	6.88
Head's highest grade completed		0.019	16.27	0.023	19.31
Head information missing	*	0.094	4.42	0.112	4.75
Hindu	*	0.109	5.11	0.119	5.38
Primary school in village	*			0.037	2.10
Primary and middle school in vill.	*			0.073	3.05
Primary, middle, and secondary in vill.	*			0.083	6.43
Nearest town within 5 km	*			0.018	1.31
Nearest railroad within 5 km	*			-0.001	-0.11
Nearest bus within 5 km	*			0.014	1.71
Paved road in village	*			0.006	0.42
Electricity in village	*			0.019	1.10
PHC clinic in village	*			-0.006	-0.27
Health subcenter in village	*			-0.011	-1.09
Hospital in village	*			-0.015	-1.00
Dispensary in village	*			0.001	0.11
Health guide in village	*			0.001	0.05
Bank in village	*			0.009	0.92
Co-op in village	*			0.007	0.55
Post-office in village	*			-0.009	-0.60
Market in village	*			-0.021	-2.95
Cinema house in village	*			0.003	0.31
Pharmacy in village	*			0.016	1.15
Mahila Mandal	*			-0.022	-1.01
Flood within the last two years	*			-0.003	-0.22
Drought in the last two years	*			-0.007	-0.56

Notes: The marginal effect for a zero/one variable is the effect of a change in the variable from zero to one on the probability of a child being in school. The specification includes dummy variables for each state, see Table 12. T-ratios refer to the underlying probit coefficient. [a] Reference group is quintile 1 (poorest). [b] Reference group is urban male.

The regressions are estimated separately for each state and then for India as a whole. The all-India regressions include dummy variables for each state (with Bihar as the reference state). Instead of presenting the complete set of equations for each of the 25

BEST COPY AVAILABLE

states and India as a whole, we first report the all-India results for each of the samples (Table 12) and then report just the wealth effects by state (Table 13). A companion paper delves more deeply into the interpretation of the other variables in the Indian context, including an examination of gender impacts and an exploration of the state specific effects (Filmer and Pritchett, 1998b).

Table 13 presents the marginal effects of being in each quintile on the probability a child aged 6 to 14 will be in school when the effects are estimated state by state in the pooled and rural only samples. While the effects are large on average, there is a substantial amount of variation across states in the magnitude of the wealth effects. For example, a child from the highest quintile in Kerala is about 5 percentage points more likely than one from the poorest quintile to be in school, whereas in Bihar the difference is 43 percentage points.[21] Focusing on rural areas only exacerbates the differences with the Kerala-Bihar difference in the wealth gap going from 4 in Kerala to 53 percentage points in Bihar.

[21] Recall that the quintiles are based on the all India sample so that the highest quintile in each state refers to the same level of wealth.

Table 13: Marginal effects of wealth on the probability of being in school for ages 6 to 14, urban and rural (Probit regression results for selected variables). States sorted by the "quintile 5" coefficient in the rural sample.

	Pooled urban and rural samples				Rural sample only			
	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Mizoram	0.030	0.073	0.112	0.083	-0.012 i	0.026 i	0.018 i	-0.096 i
Himachal Pradesh	-0.035 i	0.031 i	0.045 i	0.062 i	-0.086 i	0.005 i	0.013 i	0.026 i
Kerala	0.017 i	0.038	0.059	0.046	0.014 i	0.037	0.058	0.042
Goa	0.019 i	0.042	0.064	0.098	0.024 i	0.038	0.063	0.054
Nagaland	-0.004 i	0.027 i	0.017 i	0.064	0.001 i	0.037 i	0.007 i	0.065 i
Manipur	0.032 i	0.055	0.085	0.073	0.037	0.049	0.095	0.095
Jammu	0.039 i	0.079	0.146	0.160	0.028 i	0.066	0.118	0.119
Tamil Nadu	0.006 i	0.061	0.106	0.143	-0.001 i	0.078	0.119	0.142
Tripura	0.080	0.115	0.138	0.079 i	0.066	0.136	0.137	0.155
Delhi	0.055 i	0.072 i	0.115	0.446			0.087	0.160
Maharashtra	0.048	0.084	0.124	0.199	0.049	0.093	0.163	0.164
Assam	0.131	0.202	0.212	0.133	0.139	0.212	0.187	0.172
Haryana	0.072	0.093	0.186	0.234	0.084 i	0.107	0.229	0.196
Arunachal Pradesh	0.137	0.215	0.239	0.242	0.121	0.217	0.226	0.212
Orissa	0.082	0.206	0.231	0.263	0.095	0.229	0.250	0.251
Meghalaya	0.011 i	0.081	0.188	0.197	0.011 i	0.083 i	0.209	0.257
Gujarat	0.057	0.106	0.179	0.294	0.066	0.145	0.210	0.273
West Bengal	0.152	0.242	0.290	0.271	0.124	0.226	0.287	0.284
Punjab	0.035 i	0.104	0.207	0.336	0.022 i	0.110	0.246	0.286
Karnataka	0.088	0.185	0.253	0.296	0.074	0.191	0.267	0.303
Madhya Pradesh	0.121	0.198	0.268	0.348	0.135	0.220	0.297	0.371
Uttar Pradesh	0.135	0.188	0.271	0.382	0.152	0.196	0.282	0.372
Andhra Pradesh	0.077	0.151	0.261	0.322	0.083	0.126	0.270	0.387
Rajasthan	0.082	0.158	0.296	0.388	0.065	0.180	0.339	0.406
Bihar	0.150	0.248	0.400	0.426	0.167	0.255	0.425	0.526
All India	0.103	0.169	0.241	0.307	0.111	0.185	0.269	0.315

Source: NFHS 1992-93

Notes: All underlying probit coefficients for displayed variables are significant except those indicated by "i". Marginal effects are evaluated at the means of the other variables. In addition to the displayed variables, the probit regression includes age, age squared; gender, age, and schooling of the head of the household; a dummy for Hindu. The regression for the rural sample includes dummy variables for village infrastructure (for example for the presence of a paved road, a PHC clinic, a post office, a marketshop). All India regression includes dummy variables for state (see Table 12).

The results found here are consistent with those from other studies. For example, NCAER (1994) found that the difference in the percent of children aged 6 to 14 years old who had ever attended school between children from households with per capita incomes of less than Rs3,000 and children from households with per capita incomes of more than Rs10,000 was 25 percentage points, in an average taken over 14 major states. The range in the difference was smallest in Kerala where there was no difference found, and largest

in Punjab where it was 55 percentage points. Haque, Lanjouw and Ravallion (1998) find similar differences across the quintiles in the raw enrollment rates (see Table 7). Moreover, in a regression on enrollment in their data, the coefficient of (log) per capita consumption in explaining enrollment is 0.178. If we apply this response to the percentage difference in average consumption between the highest and lowest per capita quintiles in India, the result is very close to what we get if we use the estimates in Table 12.[22]

Last, Behrman and Knowles (1997) review estimates on the income elasticity of educational attainment from many different countries (summarized in Table 14). These are not strictly comparable as they are elasticities of attainment, not enrollment probabilities but the elasticity for the poorer countries is consistent with an estimate of close to 0.18.

Table 14: Estimates of the elasticity of schooling outcomes with respect to incomes

Country	Year	Outcome measure	Elasticity
Ghana	1987/9	School attainment	0.18-0.56*
Nepal	1980/1	Grade attainment	0.38*
Bangladesh	1980/1	Attendance	0.20
Pakistan	1989	Numeracy and literacy	0.05-0.23*
Cote d'Ivoire	1985/7	School attainment	0.14-0.42*
Bolivia	1989	Grade attained	0.04*
Nicaragua	1977/8	Grades completed	0.02-0.07
Brazil	1970	Completed years	0.09-0.16*
Brazil	1982	Completed years	0.06-0.22*
Venezuela	1987	Years	0.01*
Taiwan	1989	Years of schooling	0.12-0.33*

Source: Adapted from Behrman and Knowles (1997)
Notes: * indicates that the underlying estimate was significant at the 10 percent level. Country/years are sorted by PPP per capita GDP.

While it is beyond the scope of this paper to thoroughly explore the causes behind

[22] For example, if the difference in average per-capita consumption between the richest and poorest quintiles is 139 percent, then a marginal effect estimate of 0.178 implies a 25 (139×0.178) percentage point difference in the enrollment rates. This compares to our all India estimate of 31 percentage points difference between the poorest and richest quintiles of the wealth index.

the differences in enrollment and attainment across wealth groups, there are some important implications from this analysis. Foremost, it is clear that a simple theoretical model where household wealth and child education outcomes are unrelated is not consistent with the evidence. Such theoretical models, which generally assume that education is a pure investment, households are perfectly inter-generationally linked, the returns to education are randomly distributed across the population, and credit markets are perfect are perhaps more useful as an organizing framework for the ways reality differs from these assumptions.

The theory can break down on each of the assumptions, explored more thoroughly in Behrman and Knowles (1997). Education has a consumption (i.e. non-investment) component and richer households will consume more of it. Children from poor households face lower returns (either in reality or in perception) to schooling and hence invest smaller amounts in it. Access to credit on the basis of future potential returns to schooling may be difficult for all but especially for the poor and therefore they can finance less of it.[23]

IV) Conclusions

In this paper we show that the impact of wealth on enrollment can be estimated without income or expenditure data—without apologies or tears—using household asset variables. The use of principal components provides a set of methodologically simple yet

[23] Although credit constraints are often put forward as explaining differences in education across wealth groups, our findings cast some doubt on their importance. Since the gaps we identify are not only large but highly *variable* across states, any theory that rested on capital market imperfections would have to explain this cross-state variability in access to credit of the poorest part of the population.

defensible weights to create an index of assets which proxies for long-run wealth. In the four countries examined, India, Indonesia, Nepal, and Pakistan, this approach produces remarkably reasonable results. Education outcome differentials and wealth groups are more strongly related when the asset index, rather than a conventional consumption expenditures based measure, is used as a proxy for long-run wealth. This is consistent with there being *less* measurement error in the asset measure relative to the expenditures based measure as a proxy for long-run wealth.

The ability to generate wealth groups which are useful for analyzing educational outcomes in a consistent methodological manner using DHS-like data opens up a host of possibilities for data analysis even when income or consumption expenditures are not collected. For example, Filmer and Pritchett (1998a) explore how educational attainment profiles differ across wealth groups in the 35 countries that have had a recent DHS survey. Similar country specific (or comparative) analyses can be carried out for a wider range of socio-economic indicators included in the DHS such as health outcomes, fertility, and family planning usage for example.

When the asset index is applied to the Indian data the results show large wealth gaps in the enrollment of children which vary widely across states of India. While on average across India a rich (top 20 percent of the asset index) child is 31 percentage points more likely to be enrolled than a poor (bottom 40 percent), this ratio varies from a wealth gap of only 4.6 in Kerala, to 38.2 in Uttar Pradesh and 42.6 percentage points in Bihar.

References

Behrman, Jere R. and James C. Knowles, 1997. "How Strongly is Child Schooling Associated with Household Income?" University of Pennsylvania and Abt Associates. Mimeo.

Drèze, Jean, and P.V. Srinivasan, 1997. "Widowhood and Poverty in Rural India: Some Inferences from Household Survey Data," *Journal of Development Economics* 54:217-234.

Fields, Gary S., 1998. "Income Mobility: Meaning, Measurement, and Some Evidence for the Developing World," mimeo, Cornell University.

Filmer, Deon and Lant Pritchett, 1998a, "Education Attainment Profiles of the Poor (and Rich): DHS Evidence from Around the Globe" mimeo, DECRG and HDNED, The World Bank. Washington, DC.

Filmer, Deon and Lant Pritchett, 1998b, "Determinants of Education Enrollment in India: Child, Household, Village and State Effects," mimeo, DECRG, The World Bank. Washington, DC.

Glewwe, Paul and Gillette Hall, 1995. "Who is Most Vulnerable to Macroeconomic Shocks? : Hypotheses Tests Using Panel Data from Peru," LSMS Working Paper No. 117. The World Bank. Washington, DC.

Grosh, Margaret, and Paul Glewwe, 1998. "The World Bank's Living Standards Measurement Study Household Surveys," *Journal of Economic Perspectives*, 12:187-196.

Hammer, Jeffrey, 1998. "Health Outcomes across wealth groups in Brazil and India," mimeo, DECRG, The World Bank. Washington, DC.

Lanjouw, Peter and Martin Ravallion, 1995. "Poverty and household size," *Economic Journal: The Journal of the Royal Economic Society*. 105(433):1415-34.

Montgomery, Mark, Kathleen Burke, Edmundo Paredes, and Salman Zaidi, 1997. "Measuring Living Standards with DHS Data: Any Reason to Worry?" mimeo, Research Division, The Population Council. New York, NY.

NCAER (National Council of Applied Economic Research), 1994. "Non-enrollment, Drop-out, and Private Expenditures on Elementary Education: A Comparison across States and Population Groups." mimeo. New Delhi.

Patrinos, Harry Anthony. 1997. "Differences in Education and Earnings Across Ethnic Groups in Guatemala," *Quarterly Review of Economics and Finance* 37:809-821.

Haque, Peter Lanjouw and Martin Ravallion, 1998. "A Poverty Profile for India 1993-94," mimeo, DECRG, The World Bank. Washington, DC.

Jalan, Jyotsna and Martin Ravallion, forthcoming. "Transient Poverty in Postreform Rural China," *Journal of Comparative Economics*.

World Bank, 1997. *Primary Education in India*. The World Bank, Washington, DC.